REVIEW   **Open Access**

CrossMark

Construct and content in context: implications for language learning, teaching and assessment in China

Yan Jin (ID)

Correspondence: yjin@sjtu.edu.cn
Shanghai Jiao Tong University,
Room 2203, Haoran Hi-Tech
Building, Shanghai 200030, People's
Republic of China

Abstract

Context is vitally important in conceptualizing the construct and specifying the content of language learning, teaching, and assessment. In a rapidly changing globalized world, it is difficult but very important to identify and capture the unique features of local contexts. In this article, the experience of China will be used to discuss the impact of contextual features on policies and practices of English language education. The features of interest to the article are China's fast-growing economy in a globalized world and its recent dramatic progress in information and communications technology. To illustrate the importance of contextualized construct definition and content specification, I will use two cases to examine the alignment of contextual features with the aims and practices of English language education. In the first case, the development of the China's Standards of English shows that stakeholders' conceptualization of the construct of English language proficiency interacts with the macro-level features of the context in which activities of English language learning, teaching, and assessment are taking place. In the second case, the application of modern information and communications technology to the College English Test demonstrates the need for broadening the construct of language proficiency by adopting an interactionalist approach to construct definition, and the challenges of such an innovative approach presents for language assessment practices. The article makes the case that contextual features play a mediational function in conceptualizing and operationalizing the construct of English language proficiency and influence the policies and practices of teaching, learning, and assessment. The recognition of the role of contextual mediation in language education has important implications for language policy design and implementation in a rapidly changing world.

Keywords: Construct of language proficiency, Contextual features, English language education in China

Review

Researchers in the field of language testing and assessment have long recognized that context plays an essential role in the development and use of language assessments. Context, however, is a vague and ill-defined concept, meaning different things to different users: it can be a broad concept encompassing multiple features in the wider social context in which language assessments are developed and used, or it can refer to the specific communicative context in which language knowledge, skills, or abilities are

employed to achieve communication goals. In this section, I will undertake a review of the concept from both perspectives.

The wider social context of language assessment

Since language testers began to understand the nature of language testing as a social practice in the 1990s, the growing recognition in the field is that "the act of language testing and the language tests themselves are not neutral but are strongly embedded in political, social and educational contexts" (Shohamy 2007, p. 522). "Context" can therefore be construed as the wider social milieu in which language assessments are developed and used (e.g., McNamara 1996, 2001; McNamara and Roever 2006; Shohamy 2001). The understanding of context in this sense relates closely to the assessment use and social consequences of assessment use or the consequential aspect of validity in language assessment (Messick 1989, 1995).

In the process of developing a model for investigating and improving the impact of language assessments, Saville (2009) set out to locate impact within the complex dynamic social systems, which, in his representation, are multilayered, consisting of the general, macro context at the national level and beyond, the local-level context within the educational system (e.g., schools, communities, regions), as well as the micro-level context of individuals (e.g., teachers, learners). For a language test to achieve beneficial impact and positive washback, an examination board needs to understand "the nature of context within educational systems and the roles of stakeholders in those contexts" (Saville 2009, p. 6). The increasingly popular framework of Assessment Use Argument proposed in Bachman and Palmer (2010) is specifically designed to enable language testers to "demonstrate to stakeholders that the intended uses of their assessments are justified" (p. 2).

Context and construct in language assessment

In addition to the wider social context, the term "context" is also used in a narrower sense, primarily within the realm of construct definition in language assessment. In such cases, context is construed as factors affecting performances on language assessments, such as test methods (Bachman 1990), task characteristics (Bachman and Palmer 1996), or communicative situations in which communication tasks in a second or foreign language are to be fulfilled (e.g., Chalhoub-Deville 2003; Chapelle 1998, 1999).

Bachman (2007) provides a comprehensive summary of the strengths and weaknesses of three approaches to defining constructs in language assessment, with a focus on "the dialectic of abilities and contexts" (p. 41) (see also Purpura 2016). In a "trait/ability-focused" approach, "context" refers to the methods for eliciting language performance, or in Bachman's (1990) words, "contextual features that determine the nature of language performance that is expected for a given test or test task" (p. 112). From a "task/context-focused" perspective, the context and the task to be performed in the context are inseparable. The construct to be measured is equivalent to "ability for use" (Bachman 2007, p. 56). A strong form of the "task/context-focused" approach views construct as consisting of "abilities to accomplish particular tasks or task types" (Brown et al. 2002, p. 9, in Bachman 2007, p. 56). In the third approach, the construct of language proficiency is defined from a social interactional perspective. Context is viewed as a separate dimension, and the focus has shifted to the interaction between the ability and context.

Quoting Chapelle (1998), Bachman (2007) views an interactionalist construct definition as comprising more than trait plus context; it also includes "the metacognitive strategies (i.e., strategic competence) responsible for putting person characteristics to use in context" (p. 58).

The strongest interactionalist claim is made by He and Young (1998), who view the interaction between the context and ability as the construct; a moderate claim is presented by Chalhoub-Deville (2003), who argues that the ability and context are distinct, with the ability changing as a result of the interaction; the minimalist interactionalist claim is articulated by Chapelle (1998), who sees the ability as distinct from but interacting with the context to produce performance. In Europe, Weir (2005) advocates a socio-cognitive approach to the design and development of language assessments, which views contextual elements as an important determinant factor in establishing "context validity", that is, "the extent to which the choice of tasks in a test is representative of the larger universe of tasks of which the test is assumed to be a sample" (p. 19). The socio-cognitive approach, according to Weir (2005), "helps identify the elements of both context and processing and the relationships between them" (p. 19), and thus, it has the potential to offer a solution to the problem of generalizability when interaction between ability and context is concerned (see "Unresolved issues of an interactionalist approach to construct definition").

Aims of the study

The overarching aim of this article is to use China as an example to demonstrate how contextual features mediate the construct and content in English language education. From the brief review above, we can see that in the broad sense of the term "context," discussions center on the social consequences and impact of language assessment, whereas in the narrower sense of the term, attention is focused on the interaction between the context of language use and the ability to use the language. The article, therefore, is to achieve a twofold purpose: to discuss the influence of the contextual features in the wider society on China's policies and practices in English language education and to examine the influence of the changing context of English language communication on the conceptualization and operationalization of the construct of English language proficiency.

The question of particular interest with respect to the wider social context is how China's fast-growing economy in a globalized world has impacted its policy decisions on English language education. To address the question, I will take a case study approach drawing attention to the issues involved in the development of the China's Standards of English (CSE), a national framework of reference for English that is currently under construction. The example of the CSE is expected to showcase the impact of contextual features on the goal of the CSE, the construct of English language proficiency to be adopted in the CSE, and the complexities involved in the construction and implementation of the CSE.

From the perspective of communicative contexts, the College English Test (CET), an English language test of a very large scale and high stakes in China, will be used as a case study to argue for the need to broaden the construct of computer-mediated language communication in an era of pervasive and intense use of information and

communications technology. In the case study, I will adopt an interactionalist approach to conceptualize and understand the construct of English language proficiency. The case study will provide evidence supporting the benefits and potential pitfalls of such an approach to construct definition in language assessment.

Specifically, I will look into the impact of contextualized construct conceptualization and content specification on language education policies and practices by seeking answers to the following three questions:

1. How has China's fastest growing economy in a globalized world impacted its policies and practices of English language education?
2. How has China's rapid advancement in information and communications technology influenced the conceptualization of the construct of computer-mediated English language assessments?
3. What are the implications of "the dialectic of abilities and contexts" (Bachman 2007, p. 41) for policies and practices of English language education in China?

The main reasons for my choice of the two cases, the CSE and the CET, are their important status in the reform of English language education in China and their relevance to the discussion of the contextual mediation in making and implementing policies of English language education. In the following part, I will examine the alignment of contextual features with the aims and practices of English language education and discuss the implications of contextualized construct definitions and content specifications for English language education in China.

English in China in a global context: the case of the CSE

Half a century ago, the State Council of China proposed to establish a consistent English language education system, called "the Streamline English Language Teaching (ELT) System," for improving the efficiency and effectiveness of English language teaching, curriculum design, assessment, teacher evaluation, and training. The implementation of the proposal was suspended due to the Cultural Revolution in the 1960s and 1970s. In the 1990s, the State Council took up the proposal, but the implementation proved to be considerably more difficult than expected, due mainly to China's highly segmented educational management system (see Dai 2001; Yang 2007). The project of developing a national framework of reference for English, provisionally called China's Standards of English (CSE), was again a top-down decision made by the Ministry of Education (MOE) in the hope of improving the consistency of China's English language educational policies and practices (Lin 2015, 2016; Liu 2015). Since it was launched in 2014, the project has been managed and funded by the National Education Examinations Authority (NEEA), a governmental institution under the supervision of the MOE to take charge of educational examinations across the country. In this section, I will present the rationale for the Chinese government to launch the CSE project, analyze the roles of various stakeholder groups in the project, and discuss the complexities involved in conceptualizing and operationalizing the construct of English language proficiency arising from the dynamic interaction between the ability to use English and the context in which English is learned and used.

Setting the goal of the CSE

According to the data of the World Bank, China is the second biggest economy in the world, taking up 14.8% of the global GDP and has overtaken India as the fastest growing large economy (Suarez 2017). To remain competitive on a global scale, the Chinese government is trying increasingly to evolve a global vision in which its educational activities are seen in a wider socio-economic context. With the workplace in a globalized economy becoming a multilingual and multicultural world, the knowledge of foreign languages is a valuable asset. High proficiency in foreign languages, especially English, will enhance people's understanding and appreciation of different cultures and enable them to work in a culturally diverse environment. The development of a national framework of reference for English (i.e., the CSE), as a result, is a direct response to the Chinese government's call for a more consistent, transparent, and open English language education system that will better prepare students for the workplace in the future.

In 2013, the government proposed the "Belt and Road" (or "New Silk Road") initiative to promote the economic and cultural ties between China and its neighboring countries. Under the initiative, China is hoping to further integrate its resources, policies, and markets to connect with the outside world. Promoting exchange and cooperation in education, as a result, has become a top priority of the MOE (Ministry of Education 2017). A vital prerequisite for the attempt to expand the scope of exchange and cooperation in education is to ensure the consistency, transparency, and openness of China's education system so as to encourage the recognition of Chinese educational certificates or degrees and increase the mobility of people in and outside China.

The CSE, therefore, is intended to serve a twofold purpose (Jin et al. 2017a). Internally, the CSE is hoped to provide guidance for the reform of English curricula and English language assessments. The alignment of teaching requirements, learning objectives, and assessment criteria to the standards of the CSE will hopefully improve the consistency, coherence, and efficiency of English language education at all stages. To achieve this aim, the project team has been constructing scales of English language proficiency, which consist of categorized and calibrated descriptors of each aspect of the construct (i.e., listening, speaking, reading, writing, translation, interpretation, organizational knowledge, pragmatic knowledge) (National Education Examinations Authority 2014). Externally, the CSE is expected to establish the basis for the mutual recognition of degrees, test results, or certificates so as to promote the exchange of students or talents with neighboring countries or countries in other parts of the world. To serve this purpose, the project team has planned a series of linking studies to align the CSE levels with those of other well-established English scales or frameworks (e.g., Common European Framework of Reference for Languages, Association of Language Testers in Europe Can-Do Statements, American Council on the Teaching of Foreign Languages Guidelines, Canadian Language Benchmarks) and English language assessments (e.g., IELTS, Aptis, TOEFL iBT, PET–Academic).

### The stakeholders of the CSE and their different roles

The CSE has been constructed and will be adopted by various groups of stakeholders of English language education. As a member of the project team, I have been in charge of the construction of the CSE oral proficiency scale, through which I have gained

opportunities to come to a deeper understanding of the roles played by each stakeholder group of the project (see Fig. 1 for an illustration).

According to my understanding, there are six major groups of stakeholders, who either manage and implement the project or will use the CSE in teaching, learning, and assessment. At the top level, the MOE is the decision maker who has set the goal of the project based on its understanding of the social needs. The MOE is also expected to make policies to support the implementation of the CSE after it is put into use. The NEEA, a ministerial institution under the supervision of the MOE, has been actually managing the project and providing support of human, financial, and material resources. The NEEA is also taking charge of the validation of the CSE, at both the a priori and a posteriori stages.

The members of the CSE project team consist of experts in language testing and assessment and their master or doctoral students. This is the core group responsible for designing, constructing, and validating the CSE. The main users of the CSE include curriculum developers, textbook writers, and test developers. These groups have contributed to the construction of the CSE by participating in workshops and discussions when descriptors were developed and calibrated. At the stage of implementation, they will play a key role in revising existing curricula, textbooks, or tests or designing new ones based on the results of alignment studies. Teachers and students are also the main users of the CSE, whose role in the implementation of the CSE will be to understand and follow the requirements in the revised or new curricula, textbooks, and assessments. Other users of the CSE include educational policy-makers, employers, and admission officers, who will play a prominent role in re-considering CSE-related decisions and adjust policies if necessary.

Since the 1990s, the field of language testing and assessment has been taking a social turn which views assessment practices more as a social activity than a purely technical

Fig. 1 Major groups of stakeholders and their roles in the CSE
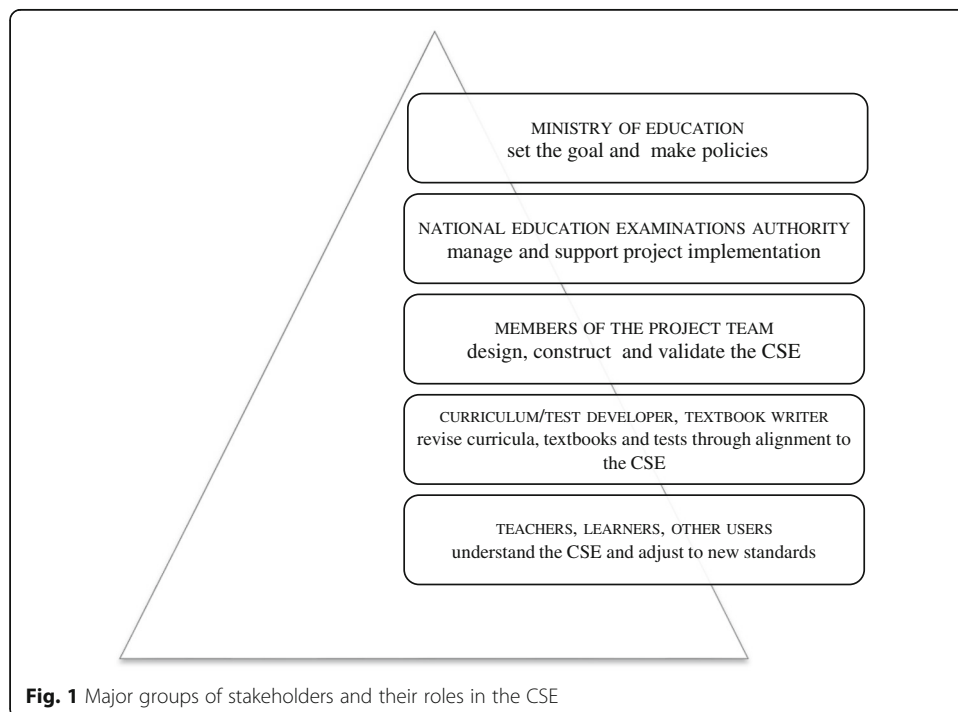
or professional activity (McNamara and Roever 2006; Shohamy 2001; Yang and Gui 2007). The CSE project is such an initiative that calls for active participation of various groups of stakeholders at each stage of its construction and implementation. The multiple groups of stakeholders and their different roles in the project have also indicated the complexity of the endeavor and the scale and depth of change which the new standards will effect on the English language education in China.

Conceptualizing and operationalizing the construct in the CSE

When the construct of English language proficiency was conceived, the project team drew on the theory of communicative competence (Canale and Swain 1980; Hymes 1972) and Bachman and Palmer's model of communicative language ability (Bachman 1990; Bachman and Palmer 1996). The team defined the construct as "the ability to use English to participate in both receptive and productive communication activities relating to certain themes and in certain communicative situations" (National Education Examinations Authority 2014). The construct was operationalized in the CSE using can-do statements about learners' knowledge of English and their use of strategies when performing six major types of communication activities, i.e., listening, speaking, reading, writing, translation, and interpretation.

Although a similar action-oriented approach has been adopted when the constructs of the CSE and the CEFR were defined, the operationalization of the constructs has differed depending on the socio-educational context in which the frameworks are developed (Jin and Jie 2017). Intended mainly for Chinese learners of English, the CSE gives priority to descriptors of learners' achievement and categorizes its descriptors based on language functions: narration, description, explanation, argumentation, instruction, and discussion (Halliday 1978; Martin and Rose 2003). The scales of the CEFR were developed primarily to serve the interests of adult language learners in Europe. As noted in Green et al. (2012), the CEFR scales "are not organised around notions and functions, but around more broadly defined competences, strategies and language activities" (p. 40). The descriptors of the CEFR "may be used both with a retrospective view towards the content of a learning programme at a given level that learners have completed (achievement) and a prospective view towards the level(s) of tasks that learners will be able to carry out beyond the classroom (proficiency)" (p. 45).

The CSE and the CEFR also differ in their structures of proficiency levels. Adopting a "branching approach," the CEFR describes finer distinctions within the three superordinate levels (A, basic; B, independent; and C, proficient) so that "the relatively small gains in language proficiency made within language programmes (achievement) can be captured and reported" (Green et al. 2012, p. 48). The CSE has hypothesized a finer-grained nine-level structure of levels, each corresponding to a key stage of English language education in China (Jin et al. 2017a, p. 13). Large-scale investigations among teachers and learners have been in progress since 2016 to empirically validate the hypothesized structure of proficiency levels.

In this section, I will use two examples, "English as a lingua franca" and "translation as a major language skill," to demonstrate how contextual features can influence decisions on the construct and content of the CSE.

Example 1: English as a lingua franca

When discussing English in Japan in the era of globalization, Seargeant (2011) noted that "(o) n various levels, and in various different ways, the English language today exhibits the trace of globalization in the forms it takes, the functions it is put to, and the attitudes that people hold towards it" (p. 2). The same holds true for English in China in the twenty-first century. One of the consequences of globalization in China is the changing status of English, which is now used more as a lingua franca than as a foreign language (Chen and Li 2017; Fan 2015). Under the circumstances, should the CSE project accept or even encourage the departure from the native-speaker norm currently adhered to in English language teaching and assessment in China? Should the teaching and testing of listening, for example, incorporate non-native English accents into instructional or assessment materials? If yes, at what level of education should "understanding non-native accents" become part of the construct of listening? Which accent(s) should Chinese learners be exposed to? With respect to speaking, should the requirement of pronunciation be "accurate" or "intelligible?" Or, should learners be penalized for speaking English with a Chinese accent? If intelligibility is the criterion, then should the accent be "intelligible" to Chinese listeners, Asians, or native-speakers?

With the rising status of English as a global language, the construct of English and its operationalization in teaching, learning, and assessment need to be reconsidered and researched. Some of the issues listed above have been considered in the construction of the CSE when proficiency levels are described, but further research is needed to support and guide the transition from the native-speaker norm to the view of English as a lingua franca. Seargeant (2011) suggests that the diversity of English in form, function, and people's attitude towards it has necessitated an analysis of the language as "situated social practice – i.e. by means of a type of almost ethnographic analysis that goes beyond a priori categories such as EFL (English as a foreign language), ESL (English as a second language) and EIL (English as an international language), and instead investigates the variegated roles played by the language in any one specific context" (p. 3). Seargeant (2011) further argues that "the processes which are referred to under the term 'globalization' do not result in uniform situations the world over, and for this reason dedicated studies of individual contexts and communities are vitally important" (p. 9).

Example 2: translation as a major language skill

Translation is a mediating language activity which occupies an important place in the normal linguistic functioning of our societies (Council of Europe 2001). Although seldom used in international language tests, translation is often employed in locally developed language tests (e.g., the College English Test in China) to assess candidates' ability to reproduce the content of a source text in the target language. There is, however, a paucity of studies on the construct of translation, i.e., the type of knowledge/skill/ability for achieving a certain level of proficiency in translation. In the current version of the CEFR, no subscales were developed to describe the level of proficiency in translation. The CSE, in contrast, has explicitly considered translation (as well as interpretation) as a major type of activity for Chinese learners of English at an intermediate or advanced level and has included it in its framework.

The importance attached to translation as a major type of language activity in the CSE partly arises from the social need for strengthening China's soft power, giving a good Chinese narrative, and better communicating China's message to the world. Since

China started its open-door policy in the late 1970s, the objective of English language education has been set as developing learners' communicative competence in English, i.e., the ability to use English as a tool of communication. The communication however was most typically one way in the 1980s and 1990s, i.e., to acquire information about what was going on outside of China. Through three decades of dramatic development, the Chinese people are now living in the fastest growing economy in the world and are highly motivated to learn English due to its role in increasing competitiveness in a globalized world. Communication in English has also become two-way, to learn about the world and to let the world learn about China. Enhancing cultural communication, for example, has been set as a goal for countries in the Belt and Road region to achieve through partnerships in education.

Cook (2007) argues that "translation has always been a useful skill, but in today's multicultural societies and globalised world it is more so than ever" (p. 398). Apart from being a means of learning, translation is also an end: "an essential skill in which one would expect the successful language learner to be competent" (Cook 2007, p. 397). He suggests that translation be added to the traditional list of four skills: reading, writing, listening, and speaking—and translating. It is further argued that apart from its role in aiding acquisition, translation has a political role to play in society: the use of translation emphasizes the equality of languages and opens people to ideas and values other than their own. By including translation in the CSE, the project team hopes to draw attention to the instruction and assessment of translation as a major language skill and contribute to China's further opening up to the outside world.

The inclusion of translation in the construct of language proficiency has nonetheless brought about an unresolved contradiction between the locality and globality of the CSE. As local standards, the CSE has every reason to consider translation as part of its construct, but as a means to improve transparency of English language education in China, the CSE needs to be aligned to international frameworks and assessments (see "Setting the goal of the CSE"), in which translation is not part of the constructs. In an increasingly globalized world, however, the status of mediation activities in language communication is changing rapidly. In a 2014–17 Council of Europe project, illustrative descriptor scales of mediation activities and mediation strategies have been developed to update the 2001 CEFR (North and Piccardo 2017). With calibrated and categorized descriptors, an alignment between the CSE levels of translation and interpretation and the CEFR levels of mediation is likely to be established. The updated version of the CEFR may also have the potential for enriching language pedagogy and assessment by encouraging the operationalization of descriptors of mediation in language teaching and assessment.

Contextualized construct definition: the case of the CET

One way to improve the consistency of China's English language education system is to align its English assessments to the standards of the CSE. By so doing, the CSE will become a contextual factor likely to have an effect on the constructs and content specifications of the assessments. Both the issues of the native-speaker norm and translation as a major language skill discussed in the previous section, for example, have been addressed in the reform of the College English Test (CET), a large-scale, high-stakes

English language assessment in China (Jin 2010, 2014; Yang 2003; Zheng and Cheng 2008). The Internet-based CET has featured non-native English accents in its listening section, and the paper-based CET has added a task of paragraph translation from Chinese into English since 2013.

The focus of this section, however, will be on the contextual feature of China's rapid advancement in information and communications technology. In recent years, English language education has been greatly enhanced by technological development, which has not only changed the way of how we communicate; more importantly, it has necessitated the conceptualization of the construct of computer-mediated or technology-enhanced language communication. In this section, I will address the second question by discussing the role of technology in the design, delivery, and scoring of the CET and stakeholders' concern over construct-irrelevant variances in the computerized or technology-facilitated CET. Examples are given to illustrate the benefits of an interactionalist perspective to language construct definition. Potential challenges presented by such a dialectic approach will also be exemplified by the case of the computer–/Internet-based CET.

The context: language communication in the digital era

More than a decade ago, Lotherington (2004) noted that "the rapid development of information and communications technology (ICT) has facilitated a revolution in how we use language" (p. 64). Over the past few years, digital literacies have been seamlessly woven into people's daily communication activities. Take WeChat in China as an example. The system was originally designed as a mobile text and voice messaging communication service. In only a few years, the service has been expanded to making audio or video calls, sending documents, and making all kinds of payments. People can now write emails, read e-books, talk via Skype or WeChat, pay via e-banking, and purchase on alibaba.com or amazon.com. Language communication in all these processes has been mediated by computers and the Internet.

The development of ICT has important implications for language testing and assessment, especially large-scale assessments targeted at tertiary-level students, a major force in the transition to computer-mediated language communication. Since the publication of the revised College English Curriculum Requirements (Higher Education Department of the Ministry of Education 2007) a decade ago, Chinese colleges and universities have been explicitly promoting the "classroom- and computer-based teaching model," which requires students to earn a certain number of credits through computer-based autonomous learning. In recent years, the advancement in artificial intelligence has further enabled human-machine collaboration in language teaching and assessment, particularly, in the area of automated scoring of open-ended speaking and writing performances and automated provision of feedback for improvement.

An interactionalist approach to construct definition

The progress in technology has necessitated a re-conceptualization of the construct of language proficiency. In the digital era, English language learners are now required to participate in digital communication as appropriately as they do in face-to-face communication. Consequently, the construct to be assessed in a language test, particularly

a computer- or Internet-based test, may be more accurately defined by taking into consideration the context in which language communication is taking place. In this section, I will use some examples of the CET to illustrate how the test construct could be more precisely defined and the test content be more carefully specified from an interactionalist perspective.

Example 1: the IB-CET writing assessment

One of the notable changes in the educational domain of the twenty-first century is that English language learners, especially those at the tertiary level of education, are now well used to writing on the computer. When the Internet-based CET (IB-CET) was introduced in 2007, the test developer was concerned with the possible influence of test mode on test takers' performances, which might confound the test construct.

As part of the validation study of the IB-CET, a comparative study was conducted to identify if there were significant differences between test takers' writing performances and processes in the paper-based CET and the IB-CET (Jin and Yan 2017). The study revealed that "scores of computer-based writing were significantly higher than those of paper-based writing, indicating a better quality of the texts produced on the computer" and that "participants, irrespective of their level of computer familiarity, made fewer language errors when writing on the computer" (p. 13). The analysis of participants' responses to the cognitive processing survey also revealed that "test takers with a higher level of computer familiarity had better perceptions of their computer-based writing processes" (p. 13).

The point made by the researchers of the comparative study was that writing on the computer may have become a norm in the digital era, and the construct of writing proficiency, accordingly, needs to be conceptualized by drawing on an internationalist view, that is, instead of being treated as an interference factor, "computer literacy should be viewed as an important contextual facet interacting with the construct measured in a computer-based language assessment" (Jin and Yan 2017, p. 1). To achieve fairness for test takers taking the test in different modes, the authors suggested that a "bias for best" approach be adopted, that is, allowing test takers to choose the test mode that fits them better (p. 16).

Example 2: the computer-based CET-SET

Similar to writing, people are now frequently engaged in non-face-to-face talking via smartphones or computers. In the computer-based CET Spoken English Test (CET-SET), such a non-face-to-face interactional task has been designed to assess test takers' ability to engage in pair discussion via computer. The focal construct to be assessed in a paired task is interactional competence (Young 2000, 2008), which, according to the model of communicative language ability, is an integral part of strategic competence (Bachman 2007; Bachman and Palmer 1996; Chapelle 1998).

An important aspect of interactional competence is the test takers' ability to use communication strategies in the process of producing a co-constructed discourse. To find out whether the mode of discussion would affect the display of test takers' strategic competence, we conducted a study to compare the use of strategies in the two modes of the CET-SET: face-to-face interview and computer-based (Jin and Zhang 2016). Adopting the method of conversation analysis, the researchers found that "interaction strategies contribute to improving the effectiveness of communication and accomplishing the communication goals in the discussion" (p. 78). More importantly, the study

revealed that the two modes of discussion share "a high degree of similarities in the quantity and variety of communication strategies," although there are "minor differences in the frequencies of cooperative strategies" (p. 75). Videos of the computer-based CET-SET show that some test takers have rich facial expressions and employ a range of body language in the discussion. Future studies need to investigate features of pair discussion in the computer-based format that are salient to raters (see May 2007, 2011).

Example 3: automated scoring of CET writing and translation

Having cited examples of assessing computer-mediated English writing or speaking, I will discuss the use of artificial intelligence (AI) in language assessment and the possible influence of the use of AI on test takers' performance. With recent progress in the development of machines with human-like intelligence in learning as well as the increase in computing power and the availability of big data (Chouard and Venema 2015), automated scoring systems have been developed and used in large-scale language assessments. The example to be used is the automated scoring of CET essay writing and paragraph translation.

In the CET writing task, test takers are required to write an essay of no less than 120 words (band 4) or 150 words (band 6) in 30 min; in the translation task, test takers are required to translate a paragraph of 140–160 Chinese characters (band 4) or 180–200 Chinese characters (band 6) into English in 30 min (National College English Testing Committee 2016). Since the CET has a test population of nine million for each administration, it takes 2 weeks for over 4000 raters in 12 marking centers across the country to complete the scoring of 9 million writing scripts and 9 million translation scripts after each test. To improve the scoring efficiency, the CET Committee has been working with an IT company on an automated scoring system.

Automated scoring poses great technical challenges due to the open-ended nature of writing and translation and even greater challenges due to the social nature of writing and translation. One of the major concerns of the CET Committee is the possible influence of automated scoring on the construct of writing or translation being assessed. The position statement of the Conference on College Composition and Communication (CCCC) in the USA states: "Writing-to-a-machine violates the essentially social nature of writing: we write to others for social purposes. If a student's first writing-experience at an institution is writing to a machine, for instance, this sends a message: writing at this institution is not valued as human communication—and this in turn reduces the validity of the assessment" (Grimes and Warschauer 2010).

With the knowledge of writing or translating to machines, test takers are likely to resort to different strategies and engage in different cognitive processes, leading to possible changes in the construct of writing or translation. Studies are therefore needed to investigate students' attitudes towards an automated scoring system and their writing and translation practices when they know that their performances are to be scored by the computer. One of the key research questions is whether test takers will "game" or exploit the weaknesses of the automated scoring system by changing their writing or translation strategies and processes when writing or translating for an inauthentic audience, i.e., machines. A study has been designed by the CET Committee in collaboration with an IT company and will be reported at the 39th Language Testing Research Colloquium (Jin et al. 2017b).

Unresolved issues of an interactionalist approach to construct definition

The case study of the CET has demonstrated the necessity and suitability of an interactionalist perspective in defining the construct in a computer-mediated or technology-enhanced language assessment. Innovative and useful as it is, "the interactional approach is not without its unresolved issues" (Bachman 2007, p.62). The main problem lies in the generalizability of test scores, i.e., performance consistencies that enable language testers to generalize across contexts. When the construct is local or co-constructed by all the participants involved, each interaction between the construct and the context is unique. Even if there was some degree of consistency in performance, it would be difficult for language testers to provide meaningful interpretations of the scores because of the complex nature of the interaction.

In Table 1, examples are provided to illustrate some unresolved issues with an interactionalist approach to construct definition in the context of computer-mediated language assessment. The difficulties and complexities lie largely in the design of computer-mediated language communication tasks and the rating and interpretation of test takers' performance on the tasks. More specifically, test developers have difficulties in operationalizing and interpreting the construct of strategic competence in computer-mediated language communication due to their lack of an in-depth understanding and clear specifications of contextual facets or interactional features.

Research of computer-mediated language assessment in the past largely focused on establishing score equivalence, rather than construct equivalence (McDonald 2002). In other words, previous empirical studies of computer-based language assessments did not have an explicit focus on the interaction between the context of language use and the ability to communicate via the computer. The study of Nakatsuhara et al. (2017) is among the few studies which examined the technology-enhanced speaking assessment on a par with the face-to-face speaking assessment. The study compared test takers' scores and linguist output as well as examiners' test administration and rating behaviors across the standard face-to-face mode and the video-conferencing mode in a high-stakes speaking test. The study however did not look into the cognitive processing of test takers, which might have important implications for the construct underlying the video-conferencing mode. Jin and Zhang (2016)

Table 1 Examples of issues with an interactionalist approach in the context of computer-mediated language assessment

	Task design and test taker performance	Rating and score reporting
Writing assessment	▪ Shall we provide test takers with auto spelling check, autocorrection, and online dictionaries when they write on the computer?	▪ How do we score test takers' performance on computer-based writing with the help of auto spelling, autocorrection, and online dictionaries? How should scores of such an assessment be interpreted and reported?
	▪ What cognitive processes are test takers expected to be engaged in when writing on the computer?	▪ How should test takers' strategic competence in a computer-based writing assessment be rated, interpreted and reported?
Speaking assessment	▪ How should test takers be paired or grouped in a computer-based pair or group discussion task?	▪ How do we score test takers who co-construct a discourse in non-face-to-face computer-mediated discussion? How should scores of such a computer-based pair discussion be interpreted and reported?
	▪ What communication strategies are expected to be used by test takers when they perform non-face-to-face computer-mediated pair or group tasks?	▪ How should test takers' strategic competence in a computer-based speaking assessment be rated, interpreted and reported?

compared the communication strategies used by test takers in the two modes of the CET-SET: face-to-face and computer-mediated. The study did not find significant differences in the quantity and variety of communication strategies in the two discussion tasks. As for computer-based writing assessments, studies have been conducted to examine test takers' cognitive processes when writing in handwritten and computer-based forms (e.g., Weir et al. 2007; Jin and Yan 2017). Both studies identified some differences in the processes of writing in different modes, but no conclusions can be drawn on the interaction between contextual features and test takers' ability to write on the computer.

When commenting on an interactionalist approach to construct definition, Bachman (2007) made a distinction among the minimalist, the moderate, and the strongest claim, depending on the degree of interaction between the context and ability that advocates of each claim would admit (see "Introduction") and clearly noted that "the research evidence in support of any of these claims, in the context of language assessment, is scanty, if not non-existent" (p. 65). Empirical studies of an interactionalist construct definition are even scarcer in the context of computer-based language assessment. Without a clear understanding of an interactionalist definition of the construct of a computer-mediated language assessment, it is highly unlikely that assessment tasks will elicit consistent performances and that raters will award reliable scores. Even a minimalist claim of an interactionalist view which sees the ability as distinct from but interacting with the context (Chapelle 1998) may present practical problems when being operationalized in language assessment. Research priority, therefore, should be given to setting up a framework of reference for contextual factors so that some degree of standardization can be achieved in terms of task design, test taker performance, and rating.

Conclusions

Bachman (2007) pointed out that "the way we view abilities and contexts – whether we see these as essentially indistinguishable or as distinct – will determine, to a large extent, the research questions we ask and how we go about investigating these empirically" (p. 41). The cases of the CSE and the CET cited in this article have further demonstrated that the way we view abilities and contexts will impact, to a large extent, the educational policies we make, how these policies will be implemented in practice, and the construct to be defined in language assessment.

The case of the CSE shows that there are multiple levels of the context in which English language education in China is taking place. The multilayered contextual mediation influences the decisions on educational policies and the implementation of the policies (see Fig. 2 for an illustration).

At the national level, the central government (e.g., the State Council) and the National People's Congress determine the needs of English language education, laying the foundation for making language educational policies. At the ministerial level, the MOE and ministerial institutions such as the NEEA set the goal of English language education based on their understanding of the social needs and formulate educational policies and plans to attain the goal. At the grass-roots level, educational institutions as well as individuals implement the policies and plans by designing and following school curricula. Contextual mediation, as indicated by the top-down arrows, determines to a large extent the constructs to be taught, learned, and assessed. The implementation of curricula and assessments, on the other hand, would bring about the so-called

Fig. 2 English language education in China: the multilayered context
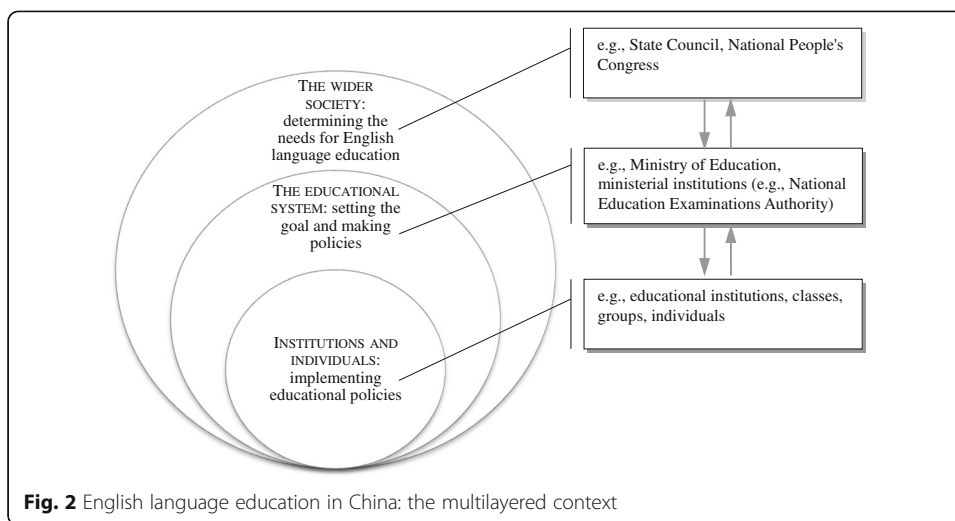
washback and impact on the education system and the society, as shown by the bottom-up arrows. The dynamic interaction between the context and the policies and practices of English language education poses the greatest and potentially most rewarding challenge to professionals working in the area of social science.

The complexities involved in an interactionalist view of the construct of computerized language assessment are also evidenced in the case of the CET. To operationalize the construct of computer-mediated language communication, we need a clearer conceptualization of the context in which assessments are constructed and a better understanding of the mediational function of context on assessment design and implementation. More importantly, we need to link a contextualized construct conceptualization to validity arguments. In other words, we should fully incorporate "context validity" (Weir 2005) into considerations of assessment development and use. By so doing, we will be able to anticipate, rather than predict, the positive and desirable impact of language assessment on policies and practices of language teaching and learning, achieving the goal of "impact by design" (Salamoura et al. 2014; Saville 2009, 2012).

Messick's (1989, 1995) view of validity stresses the need for test constructs to be relevant and useful in the testing context and for consequences of test use to be beneficial to society. In China, English language assessments are used extensively in society for a variety of high-stakes purposes, from admission to graduation, from employment to promotion, and from civil service to residential permits (Cheng and Curtis 2010; Jin 2008, 2014; Yu and Jin 2016). Messick's view of validity, according to McNamara (2001), emphasizes "the social and socially constructed nature of assessment" (p. 334) and is therefore highly relevant to English language assessment in the Chinese context. As test constructs can be seen as "the embodiment of social values," our conceptualization of test construct and specification of test content should be contextualized so as to "engage explicitly with the fundamentally social character of assessment at every point" (McNamara 2001, p. 336). Bachman (1990) has rightly pointed out about three decades ago that "tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large" (p. 279). In my view, contextualized construct definition, whether in the broad sense of the social context or in the narrower sense of

the communicative context, presents a more accurate and comprehensive picture of the nature of language proficiency and provides useful guidance on educational policy-making and practices.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–71). Ontario, CA: University of Ottawa Press.
Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
Brown, J. D., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments. SLTCC technical report 24*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i at Manoa.
Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47.
Chalhoub-Deville, M. (2003). Second language interaction: current perspectives and future trends. *Language Testing, 20*(4), 369–383.
Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). New York: Cambridge University Press.
Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics, 19*, 254–272.
Chen, X., & Li, J. (2017). On the development of intercultural communication competence in the context of English as a lingua franca. *Contemporary Foreign Languages Studies, 1*, 19–24.
Cheng, L., & Curti, A. (Eds.). (2010). *English language assessment and the Chinese learner*. Taylor & Francis Group: Routledge.
Chouard, T., & Venema, L. (2015). Machine intelligence. *Nature, 521*(7553), 435–435.
Council of Europe. (2001). *Council of Europe. Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
Cook, G. (2007). A thing of the future: translation in language learning. *International Journal of Applied Linguistics, 17*(3), 396–401.
Dai, W. (2001). The construction of the streamline ELT system in China. *Foreign Language Teaching and Research, 33*(5), 322–327.
Fan, Y. (2015). The globalization and localization of English from the perspective of English as a lingua franca and implications for "China English" and English language education in China. *Contemporary Foreign Languages Studies, 6*, 29–33.
Green, A., Trim, J., & Hawkey, R. (2012). *Language functions revisited: theoretical and empirical bases for language construct definition across the ability range*. Cambridge: Cambridge University Press.
Grimes, D. & Warschauer, M. (2010). Utility in a fallible Tool: a multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6). Retrieved February 20, 2017 from http://www.jtla.org .
Halliday, M. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: Arnold.
He, A., & Young, R. (1998). Language proficiency interviews: a discourse approach. In R. Young & A. He (Eds.), *Talking and testing* (pp. 1–24). Amsterdam: John Benjamins.
Higher Education Department of the Ministry of Education. (2007). *College English curriculum requirements*. Shanghai: Shanghai Foreign Language Education Press.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth: Penguin.

Jin, Y. (2008). Powerful tests, powerless test designers? Challenges facing the College English Test. *English Language Teaching in China, 31*(5), 3–11.

Jin, Y. (2010). The National College English Testing Committee. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 44–59). Taylor & Francis Group: Routledge.

Jin, Y. (2014). The limits of language tests and language testing: challenges and opportunities facing the college English test. In Coniam, D. (ed.). *English Language Education and Assessment: Recent Developments in Hong Kong and the Chinese Mainland* (pp. 155–169). Singapore: Springer.

Jin, Y., & Jie, W. (2017). Principles and methods of developing the speaking scale of the China standards of English. *Foreign Language World, 179*(2), 10–19.

Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017a). Developing the China standards of English: challenges at macropolitical and micropolitical levels. *Language Testing in Asia, 7*(1), 1–19. https://doi.org/10.1186/s40468-017-0032-5 .

Jin Y. & Yan, M. (2017). Computer literacy and the construct validity of a highstakes computer-based writing assessment. *Language Assessment Quarterly, 14*(2): 101-119.

Jin, Y., & Zhang, L. (2016). The impact of test mode on the use of communication strategies in paired discussion. In G. Yu & Y. Jin (Eds.), *Assessing Chinese learners of English: language constructs, consequences and conundrums* (pp. 61–84). London: Palgrave Macmillan.

Jin, Y., Zhu, B., & Wang, W. (2017b). Writing to the machine: challenges facing automated scoring in the College English Test in China. Paper to be presented at the symposium "*Human-machine teaming up for language assessment: The need for extending the scope of assessment literacy*", 39th Language Testing Research Colloquium, July 16–22, Bogota, Colombia.

Lin, H. (2015). On the reform of national matriculation examinations and the development of a national foreign language assessment system. *China Examinations, 1*, 3–6.

Lin, H. (2016). Developing a national foreign language assessment system and improving Chinese people's language proficiency. *China Examinations, 12*, 3–4.

Liu, J. (2015). The fundamental considerations in developing a national scale of English. *China Examinations, 1*, 8–11.

Lotherington, H. (2004). What four skills? Redefining language and literacy standards for ELT in the digital era. *TESL Canada Journal, 22*(1), 64–78.

Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.

May, L. (2007). *Interaction in a paired speaking test: the rater's perspective*. Unpublished PhD dissertation, University of Melbourne, Australia.

May, L. (2011). Interactional competence in a paired speaking test: features salient to raters. *Language Assessment Quarterly, 8*(2), 127–145.

McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education, 39*(3), 299–312.

McNamara, T. (1996). *Language testing*. Oxford: Oxford University Press.

McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing, 18*(4), 333–349.

McNamara, T., & Roever, C. (2006). *Language testing: the social dimension*. Malden, MA: Blackwell Publishing.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment. *The American Psychologist, 50*(9), 741–749.

Ministry of Education. (2017). Top priorities of the year plan of the Ministry of Education in 2017. [MOE Document, 2017 No. 4] retrieved on February 15, 2017 at http://www.moe.gov.cn/srcsite/A02/s7049/201702/t20170214_296174.html

Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2017). Exploring the use of video-conferencing technology in the assessment of spoken language: a mixed-methods study. *Language Assessment Quarterly, 14*(1), 1–18.

National College English Testing Committee. (2016). *Syllabus for the College English Test (CET)*. Shanghai: Shanghai Jiao Tong University Press.

National Education Examinations Authority. (2014). *A working plan for the construction of a national scale of English*. Unpublished document drafted and used by the CSE project team, Beijing, China.

North, B. & Piccardo, E. (2017). *Mediation and exploiting one's plurilingual repertoire: exploring classroom potential with proposed new CEFR descriptors*. Workshop at the 6th ALTE International Conference, May 3–5, Bologna, Italy.

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal, 100*(S1), 190–208.

Salamoura, A., Khalifa, H., & Docherty, C. (2014). *Investigating the impact of language tests in their educational context*. IAEA 2014 paper, Cambridge English Language Assessment.

Saville, N. (2009). *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider*. Unpublished PhD thesis, University of Bedfordshire, UK.

Saville, N. (2012). Applying a model for investigating the impact of language assessment within educational contexts: the Cambridge ESOL approach. *Research Notes, 50*, 4–8.

Seargeant, P. (Ed.). (2011). *English in Japan in the era of globalization*. London: Palgrave Macmillan.

Shohamy, E. (2001). *The power of tests: a critical perspective on the uses of language tests*. Harlow, England: Longman.

Shohamy, E. (2007). The power of language tests, the power of the English language and the role of ELT. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (Vol. 15, pp. 521–531). New York: Springer.

Suarez, E. (2017). The world's 10 biggest economies in 2017. Published at Linkedin.com on March 17, 2017. Retrieved on April 25, 2017, at www.linkedin.com/pulse/worlds-10-biggest-economies-2017-enrique-suarez/

Weir, C. J. (2005). *Language testing and validation: an evidence-based approach*. New York: Palgrave Macmillan.

Weir, C. J., O'Sullivan, B., Jin, Y., & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS writing component: effects and impact. *IELTS Research Reports, 7*, 311–347.

Yang, H. (2003). Fifteen years of the College English Test. *Journal of Foreign Languages, 3*, 21–29.

Yang, H., & Gui, S. (2007). Developing a common Asian framework of reference for English. *Foreign Languages in China, 2*, 34–37.

Yang, P. (2007). The construction of the streamline English teaching and material development system. Reading, Writing and Calculating: Quality Education Forum, 04X: 193–194. Retrieved on June 2, 2017 at http://gbjc.bnup.com/news.php?id=13219.

Young, R. F. (2000). *Interactional competence: challenges for validity*. Paper presented at the American Association for Applied Linguistics, Vancouver, BC.

Young, R. F. (2008). *Language and interaction*. New York: Routledge.

Yu, G. & Jin, Y. (eds.) (2016). *Assessing Chinese learners of English: language constructs, consequences and conundrums*. London: Palgrave Macmillan.

Zheng, Y., & Cheng, L. (2008). The College English Test (CET) in China. *Language Testing, 25*(3), 408–417.